MEDICAL CARE
Volume 37, Number 11, pp 1128–1139


# The Performance of Different Lookback Periods and Sources of Information for Charlson Comorbidity Adjustment in Medicare Claims

JAMES X. ZHANG, PHD,* THEODORE J. IWASHYNA, AB,†‡§
AND NICHOLAS A. CHRISTAKIS, MD, PHD, MPH*§||

**BACKGROUND. The Charlson Score is a particularly popular form of comorbidity adjustment in claims data analysis. However, the effects of certain implementation decisions have not been empirically examined.**

**OBJECTIVE. To determine the effects of alternative data sources and lookback periods on the performance of Charlson scores in the prediction of mortality following hospitalization.**

**SUBJECTS. A representative sample of 1,387 elderly patients hospitalized in 1993, drawn from the Medicare Current Beneficiary Survey (MCBS). Three years of linked Medicare claims and survey instruments were available for all patients, as was 2-year mortality follow-up.**

**STATISTICAL METHODS. Nested Cox regression and comparisons of areas under the Receiver Operating Characteristic (ROC) curve were used to evaluate ability to predict mortality.**

**RESULTS. Compared with a 1-year lookback involving solely inpatient claims, statistically and empirically significant improvements in the prediction of mortality are obtained by incorporating alternative sources of data (particularly 2 years of inpatient data and 1 year of outpatient and auxiliary claims), but only if indices derived from distinct sources of data are entered into the regression distinctly. The area under the ROC curve for 1-year mortality predication increases from 0.702 to 0.741 ($P$ = 0.002). Furthermore, these improvements in explanatory power obtained whether one also controls for Charlson scores based on self-reported health history and/or secondary diagnoses from the claim for the index hospitalization itself. Finally, claims-based comorbidity adjustment performs comparably to survey-derived adjustment, with areas under the ROC curve of 0.702 and 0.704, respectively.**

**CONCLUSIONS. The widespread practice of comorbidity adjustment in pre-existing administrative data sources can be improved by taking more complete advantage of existing administrative data sources.**

**Key Words: Medicare; co-morbidity; data quality. (Med Care 1999;37:1128–1139)**

Among the most popular comorbidity indices in claims data research are those based on the work of Mary Charlson et al,[1] particularly as implemented in the International Classification of Dis-

*From the Department of Medicine, the University of Chicago, Chicago, Illinois.

†From the Harris School of Public Policy, the University of Chicago, Chicago, Illinois.

‡From the Pritzker School of Medicine, the University of Chicago, Chicago, Illinois.

§From the Population Research Center, the University of Chicago, Chicago, Illinois.

||From the Department of Sociology at the University of Chicago, Chicago, Illinois.

This work was supported by a grant from the Alzheimer's Association (TRG-95-033), the National Institute on Aging (1 R01 AG15326-01) (NAC), and by a Medical Scientist National Research Service Award from the National Institutes of Health (5 T32 GM07281) (TJI).

Address correspondence to: Nicholas A. Christakis, MD, PhD, MPH, Section of General Internal Medicine, The University of Chicago Medical Center, 5841 S. Maryland Ave., MC 2007, Chicago, IL 60637. E-mail: nchrista@medicine.bsd.uchicago.edu

Received October 21, 1998; initial review completed December 30, 1998; accepted May 25, 1999.

eases, 9th Revision, Clinical Modification (ICD-9-CM) codes for computerized use.[2–4] While several alternative risk-adjustment approaches have also been published,[5–9] the Charlson method is extremely popular.[4,10–13] Direct comparisons between different comorbidity measures are relatively rare, however.[14–17] In general, these indices have been developed to predict mortality following hospitalization (a pattern to which we will adhere), although alternative outcomes[2,14,18,19] and settings[20–25] have also been evaluated.

In implementing comorbidity adjustment for mortality risk following hospitalization, many practical issues have been decided on the basis of convenience, experience, judgement, and data availability, rather than on an empirical examination of the effects of these decisions on the performance of the comorbidity index in question. Two areas have received considerable attention: (1) the difficulties caused by the disease coding schemes used in administrative databases[26–28] and related data-quality issues,[15,29–31] and (2) the question of study-specific reweighting of the Charlson Index.[14,17,32] However, a number of other important issues are only beginning to be addressed, particularly related to the amount of longitudinal data that should be collected on each individual. Thus, researchers working with California's Office of Statewide Health Planning and Development discharge abstract data regularly perform risk adjustment without the ability to link to any previous claims,[27] whereas those using Medicare claims data may utilize all inpatient claims for several years.[2,4,13] Similarly, in the development of incidence cohorts, the appropriate "lookback" time (that is, the amount of retrospective surveillance necessary to ensure that the disease is incident and not prevalent) has been carefully examined in cancer,[33] although not in other diseases and not for risk-adjustment purposes. Study has also begun on the value of administrative sources of data other than inpatient claims (eg, with respect to cancer incidence,[34] comorbidity adjustment,[35] or the use of supplementary survey-derived data[25,36–38]).

Here, we take advantage of the longitudinal, individually linked inpatient, outpatient, and physician claims available in the Medicare Current Beneficiary Survey. For a cohort of patients hospitalized in 1993, we examine the impact of alternative lengths of lookback (1 vs. 2 years) and of alternative data sources (inpatient claims only vs. inpatient plus "outpatient" and "auxiliary" claims vs. all these claims plus self-report) on the performance of the Charlson score with respect to mortality following hospital admission. We specifically evaluate two different ways the comorbidity information may be combined statistically; that is, we evaluate whether it makes a difference if the data from different sources (eg, inpatient and outpatient claims) are combined into a single overall Charlson score (the usual approach), or, alternatively, are kept distinct, with separate Charlson scores developed for each data source and entered into regression models as distinct vectors of covariates.

Our general approach is to assume that most researchers study patients initially identified in inpatient claims. We consider the marginal value of additional sources of other earlier claims-based information once one already has controlled for comorbidity levels detected in the earlier inpatient claims. This imposes an a priori hierarchy on the claims, looking first at inpatient claims, second at outpatient claims, and finally at auxiliary claims. We also consider the value of two other sources of information to supplement inpatient claims-based lookbacks: the use of self-reported data on medical history and secondary diagnoses present on the claim for the index hospitalization.

## Methods

### Sources of Data

The cohort was drawn from the 1991 cohort of the MCBS. This nationally-representative sample of approximately 13,000 Medicare beneficiaries drawn in 1991 continues to be maintained, as described elsewhere.[39] The MCBS contains quarterly survey data linked to all Medicare claims (including inpatient, outpatient, and auxiliary service claims) filed during the calendar year during which the subject is followed in the survey. We have developed a panel data set by linking the releases of the MCBS from 1991 through 1994.

Our study cohort consisted of all individuals aged 65 years and older in the original MCBS sample (in 1991) who were hospitalized in 1993. Subjects entered the study cohort upon admission to the hospital for the first time in the 1993 calendar year, their "index hospitalization." For those hospitalized more than once in 1993, subsequent hospitalizations were ignored. We required that cohort members be at least 67 years

old in 1993 to allow for 2 years of previous Medicare claims to be inspected for the development of the comorbidity indicators. In- and out-of-hospital mortality follow-up was available through January 1, 1995, for all cohort members, at which point survival was censored.

## Construction of Charlson Comorbidity Measures

**Claims-Based Algorithm.** For each time period (eg, a 1-year lookback and a second-year lookback), a comorbidity score was generated for each cohort member by searching through the entire MCBS Medicare files; these include "inpatient" claims (hospital), "outpatient" claims (which in HCFA terminology are claims for outpatient care filed by *institutional* providers), and "auxiliary" service claims (physician/supplier, skilled nursing facility, home health aid, hospice, and durable medical equipment). Traditional, office-based outpatient care is typically billed in the physician/supplier claims; however, tests indicated that distinguishing physician/supplier claims from other auxiliary claims did not increase the explanatory power of the models (data not shown).

The algorithm we used to search the claims and to assign Charlson scores is a minor variant of the Deyo[2] and Romano[3] methods; in particular, we employed ICD-9-CM condition codes appearing in either method, but excluded the procedure codes advocated by Romano. Two lookback periods were established: a 1-year lookback (days, 1–365) and a second prior-year lookback (days, 366–730). Day 1 for the lookback periods is the day preceding admission for the index hospitalization. The following abbreviations are defined in Table 1 and used in the tables and figures to clarify the way Charlson measures are constructed. "In (1)" is a Charlson score based on the inpatient claims from the 365 days preceding the index hospitalization. "In (2)" is a Charlson score based on the diagnoses present on inpatient claims from only the second prior year of inpatient claims, regardless of the diagnoses present in the first year or in any other data source. Likewise, "Out (1)" is the Charlson score based on a 1-year lookback in the institutional outpatient claims, and "Aux (1)" is the Charlson score based on a 1-year lookback in all other claims, the so-called "auxiliary" claims. We also constructed a Charlson score based on the secondary diagnoses (up to nine are recorded in the claims) from the index hospitalization ("Sec").

**Self-Report-Based Algorithm.** The MCBS contains questions allowing for a self-reported history of the following diseases: "hardening of the arteries," myocardial infarction, angina, stroke, brain hemorrhage, cancer, diabetes, rheumatoid arthritis, Alzheimer's disease, emphysema, asthma, chronic obstructive pulmonary disease, and partial paralysis. Some of these questions correspond to certain Charlson score categories, and points were assigned as appropriate. Because severity could not be determined from the MCBS questionnaire, diseases with multiple severity levels in the Charlson system were assigned to the lowest severity level.

## Statistical Analyses

**Modeling Methodologies.** Cox regression was used to model the effects of alternative comorbidity measures. All regressions control for race (white vs. non-white), gender, and age (captured as age and age squared). Furthermore, the primary diagnoses of the index hospitalization were categorized into 18 categories, which were formed for consistency with previous typologies[40] and to ensure that no one group was too small (the results were not sensitive to the particular categorization employed for the index hospitalization primary diagnosis [data not shown]). These diagnostic categories were treated as "nuisance parameters" in the estimation of the Cox models, allowing for maximal flexibility without requiring proportionality in the shape of the hazard function across disease categories[41]; in doing so, however, separate coefficients are not estimated for these variables. Selected Cox-regression likelihood ratio $\chi^2$ statistics are presented and are denoted as $G^2$. The likelihood ratio $\chi^2$ statistics can be converted into an $R^2$ analog by using the formula

$$R^2 = 1 - \exp(-G^2/n)$$

in which $n$ = 1,387 here.[42] For nested Cox models, the difference in the $G^2$ between models, denoted $\Delta G^2$, has a $\chi^2$ distribution with as many degrees of freedom as there are differences in the number of covariates between the models.

Presented ROC curves are based on the prediction of 365-day mortality following admission using the same covariates as the Cox regression.[43] Statistical comparisons were performed using

TABLE 1. Simple Statistics on Select Comorbidity Measures

| Data Source | *n* | Mean | SD | Maximum |
|---|---|---|---|---|
| Year-1 inpatient lookback ["*In (1)*"] | 282 | 1.44 | 1.58 | 8 |
| Inpatient lookback for year-2 ("*In (2)*"]* | 402 | 1.46 | 1.49 | 7 |
| Year-1 outpatient lookback ["*Out (1)*"] | 970 | 0.71 | 0.71 | 8 |
| Year-1 auxiliary lookback ["*Aux (1)*"] | 1378 | 1.36 | 1.36 | 10 |
| In (1) + In (2) + Out (1) + Aux (1) | 1378 | 1.77 | 1.93 | 12 |
| Self-reported disease history ["*Self*"] | 1385 | 1.42 | 1.34 | 9 |
| 2° Diagnoses from index hospitalization ["*Sec*"] | 1387 | 0.89 | 1.21 | 8 |
| Maximum: In (1) + In (2) + Out (1) + Aux (1) + Sec + Self | 1387 | 2.91 | 2.20 | 14 |

Note: This table contains the number of individuals for whom each Charlson score was observed, and the mean, standard deviation, and maximum Charlson score realized in the claims of those patients who had at least one claim in the respective data sources.

*This is a Charlson score based on the 730th through the 366th day before admission.

ROCKIT, (University of Chicago, Department of Radiology. Chicago, Illinois. www-radiology.uchicago.edu/sections/roc).[44,45] Probabilities are reported for one-sided comparisons for the statistically significant increase in the area under the ROC curve. The conventional probability levels of significance ($P \leq 0.1$ worthy of report; $P \leq 0.05$ as significant) were used.

**Interpretation of Statistical Tests.** Comparisons of three types are made across models. ROC analysis is used to compare models that contain different (non-nested) sets of variables based on logistic regression predicting mortality within 365 days of hospitalization. This has the virtue of easy comparability across models and familiarity. However, logistic regression cannot make full use of the detail of the mortality data that is available; thus, we also use Cox regression to capture all the information about when patients die. For two distinct purposes, we use $G^2$ and an $R^2$ analog when examining Cox models. We use $G^2$ to allow formal statistical comparisons of nested models (ie, comparisons between two regressions in which the covariates of one model are a subset of the other model). This is analogous to the use of F-tests in ordinary least-squares regression. To compare non-nested models, we provide $R^2$ analogs, which, while often appearing trivially small for Cox regression models, nevertheless allow the comparison of relative magnitudes. In summary, while each indicator is imperfect in some way, we use triangulation across all three to present the best-supported analysis of the data.

**Parameterization of the Charlson Score: Indicator Variables Used.** In all cases, an indicator-variable approach was taken when including the Charlson score in regressions, as has been suggested elsewhere.[10] In practice, this means that a set of dummy variables was constructed for each patient for each Charlson score value; if their Charlson score was equal to 2, then the dummy for "Charlson is 2" was set to 1, and all others (eg, the dummies "Charlson is unobserved," "Charlson is observed to be zero," "Charlson is 1," "Charlson is 3" "Charlson is 4 or greater") were set to zero. Two differences with previous work are important to note. First, we distinguished between individuals without any claim filed during the lookback window ("unobserved Charlson"), and those for whom at least one claim was filed but on which no Charlson diseases were indicated ("observed Charlson of zero"). In past work, these groups often appear to be combined and assigned a Charlson value of zero. Second, because of the relatively small number of individuals who had Charlson scores of four and greater, these higher values were combined into a single category.

We also tested a linear, continuous specification of the Charlson score and found with one exception the same patterns reported later. With the linear specification, the second year of inpatient data appears to be less valuable than with the specification employed here (data not shown).

**Parameterization of the Charlson Score: "Single" Versus "Separate" Vectors.** Finally, we tested two alternative ways to incorporate

alternative sources of data and lookback periods. In the first method, a model was specified which combined all data sources into a single Charlson score without regard to the data source in which a constituent disease was detected. In such models, there were a total of five variables indicating levels of the Charlson score, as explained earlier. Thus, Cox regression models took the form:

$$\ln h = \beta_1 \cdot \mathbf{Dx} + \beta_2 \cdot \mathbf{Dem} + \beta_3 \cdot \mathbf{C}$$

in which **Dx** is a vector of index hospitalization primary-diagnosis indicator variables treated as a nuisance parameter (so $\beta_1$ is not explicitly estimated), **Dem** is a vector of demographics variables, and **C** is a set of five indicator variables for the levels of the Charlson score. This is the "single vector" Charlson specification.

An alternative approach allows separate Charlson scores based on each data source and enters them into the regression separately, as in:

$$\ln h = \beta_1 \cdot \mathbf{Dx} + \beta_2 \cdot \mathbf{Dem} + \beta_3 \cdot \mathbf{C_{in}} + \beta_4 \cdot \mathbf{C_{out}} + \beta_5 \cdot \mathbf{C_{aux}}$$

in which $\mathbf{C_{in}}$ is a vector of five indicator variables for the level of a Charlson score based on inpatient data, $\mathbf{C_{out}}$ is a vector of five indicator variables for the level of a Charlson score based on outpatient data, and $\mathbf{C_{aux}}$ is a vector of five indicator variables for the level of a Charlson score based on auxiliary claims data. This is the "separate vector" Charlson score specification. Note that, in this specification, a single Charlson-diagnosis (eg, chronic obstructive pulmonary disease) could contribute to the score of both $\mathbf{C_{in}}$ and $\mathbf{C_{out}}$ if it was noted in both the inpatient and outpatient claims.

## Results

### Descriptive Statistics of Study Cohort

Of the elderly subjects (> 67 years in 1993) in the MCBS, 1,387 were hospitalized at least once in 1993. Their mean age was 78.2 years (standard deviation: ± 7.7) and 38.8% were male, 86.7% were white, and 158 (11.4%) had died by January 1, 1995. The mean Charlson score assigned to those who had a claim of each type is presented in Table 1 for each data source. In general, scores developed from distinct data sources have moderate correlations, typically in the 0.25 to 0.50 range (data not shown).

### "Single Vector" Charlson: Negligible Value of Additional Data Sources

As shown in Fig. 1 and Table 2, when data from different sources are combined into a *single* Charlson index, there is no clearly superior combination of data sources. A single year of inpatient data performs as well as a Charlson index based on any combination of inpatient, outpatient, and auxiliary claims. All provide an area under the ROC curve of approximately 0.70 for 1-year mortality, and nearly identical $R^2$.

Moreover, for these single-index Charlson scores, the areas under the ROC curve are only minimally different from the area obtained by simply adjusting for age, race, sex, and primary diagnosis of the index hospitalization (Table 2, "Single Vector" column). The areas under the ROC curve are not statistically different between models with and without a single-vector Charlson score at conventional levels. The $G^2$ statistic indicates that the inpatient Charlson scores do significantly increase the explanatory power of the Cox model (Fig. 1, $\Delta G^2 = 53.1 - 42.5 = 10.6$, 5 d.f., $P = 0.06$). However, there is no particular advantage to any multiple-data source, single-vector Charlson relative to inpatient-only single vector Charlson scores (Fig. 1).

### "Separate Vector" Charlson: Complementarity of Alternative Sources of Claims Data and Longer Inpatient Lookbacks

An alternative approach to judging the value of the different sources of information is to create separate Charlson indices from each data source and to enter them into regressions separately. This amounts to acknowledging that diseases recorded in inpatient claims and diseases recorded in other claims may have different import with respect to their severity and, hence, should be allowed to have a different impact on mortality. A similar argument might be made about time horizon (eg, with more recently detected diseases in still-living individuals being more "severe" than long-standing ones). This more flexible parameteriza-

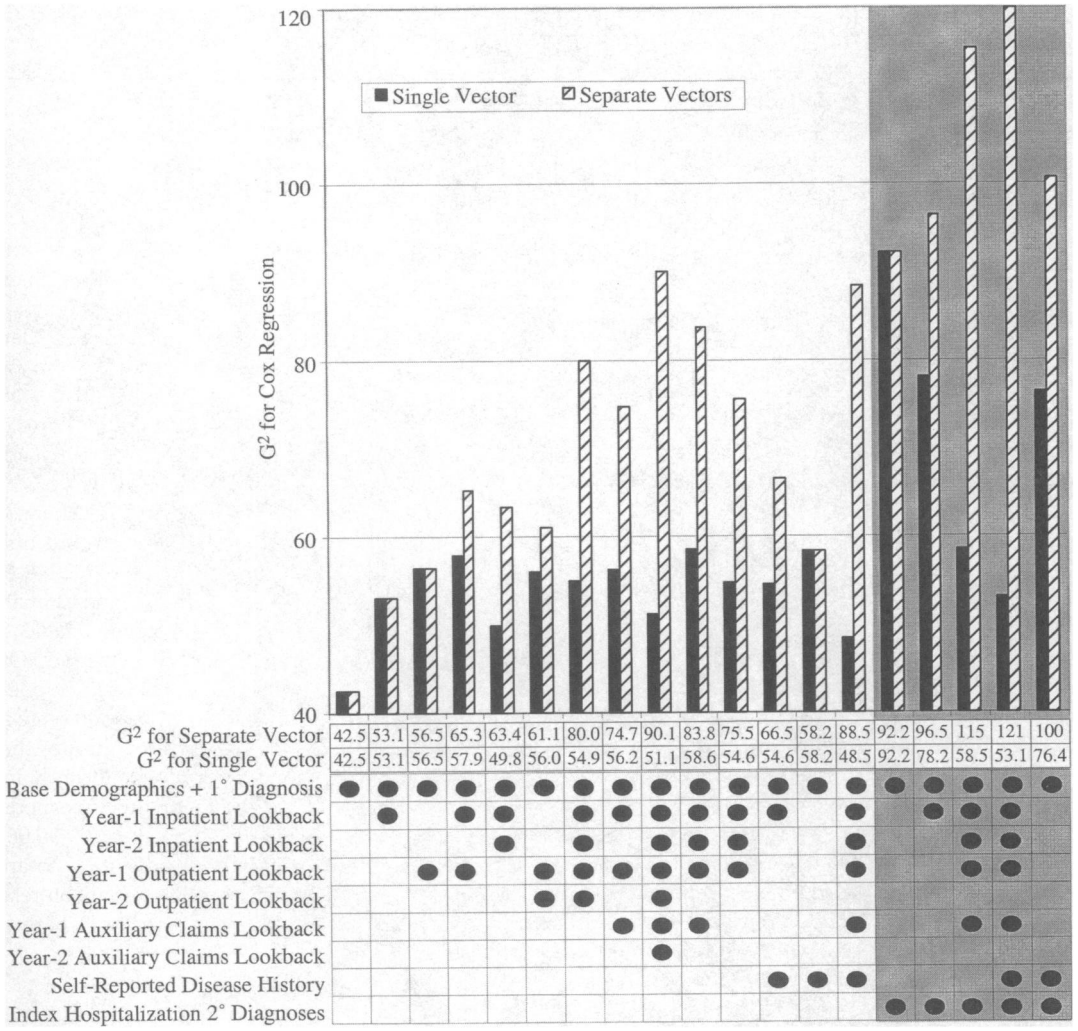


FIG. 1. Likelihood ratio $\chi^2$ statistics ($G^2$) from Cox regressions with alternative comorbidity measures. These likelihood ratios ($G^2$) for the alternative comorbidity measures show the increase in explanatory power associated with different data sources and parameterizations. When comparing nested models for which the Charlson scores were entered as separate vectors, the difference in the $G^2$ is $\chi^2$ distributed; each data source provides five degrees of freedom. Non-nested models cannot be directly compared using $G^2$; $G^2$ must be converted to $R^2$ analogs using the formula $R^2=1-\exp(-G^2/1387)$. Note that the $R^2$ analogs are a monotonic function of $G^2$ scores. All models presented control for the age, race, gender, ("base demographics") and primary diagnosis on index hospitalization of the patients. Models including Charlson scores based on secondary diagnoses from the index hospitalization are distinguished in the figure by the gray background on the right, as there are important conceptual difficulties in the interpretation of these data as comorbidities.

tion, taking advantage of the implicit information in data source, reveals that alternative sources of information do have some value.

As reported earlier, there was a 10.6-point increase in $G^2$ associated with the addition of the 1-year inpatient lookback to a model which only controlled for demographics and index-hospitalization primary diagnosis (Fig. 1, $\Delta G^2$ = 53.1 - 42.5 = 10.6, 5 d.f., $P$ = 0.06). Using separate vectors, each alternative source of Charlson scores within the 1-year lookback appears to be detecting important and different comorbidity. Thus, there is

TABLE 2. Performance of Alternative Data Sources as Measured by the Area Under the ROC Curve

| | Single Vector | Separate Vectors |
|---|---|---|
| Base Demographics + 1° Diagnosis [*No Charlson Score*] | 0.697 | 0.697 |
| Year-1 Inpatient Lookback [*"In (1)"*] | 0.702 | 0.702 |
| Year-1 Inpatient + Year-1 Outpatient Lookback [*"In (1) + Out (1)"*] | 0.710 | 0.719 |
| Year-1 Inpatient + Year-1 Outpatient + Year-1 Auxiliary Claims Lookback [*"In (1) + Out (1) + Aux (1)"*] | 0.708 | 0.724 |
| Year-1 Inpatient + Year-2 Inpatient Lookback [*"In (1) + In (2)"*] | 0.697 | 0.720 |
| In (1) + In (2) + Out (1) + Out (2) | 0.710 | 0.730 |
| In (1) + In (2) + Out (1) + Out (2) + Aux (1) + Aux (2) | 0.702 | 0.733 |
| In (1) + Out (1) + Aux (1) + In (2) | 0.710 | 0.741 |
| In (1) + Out (1) + In (2) | 0.705 | 0.727 |
| Self-Reported Disease History [*"Self"*] | 0.704 | 0.704 |
| In (1) + Self | 0.700 | 0.713 |
| In (1) + Out (1) + Aux (1) + In (2) + Self | 0.703 | 0.743 |
| 2° Diagnoses from Index Hospitalization [*"Sec"*] + In (1) | 0.718 | 0.727 |
| Sec + In (1) + Out (1) + Aux (1) + In (2) | 0.704 | 0.751 |

Note: All models which include Charlson scores also control for the base demographics and index hospitalization primary diagnosis.

an increase of 12.2 points (Fig. 1, $\Delta G^2 = 65.3 - 53.1 = 12.2$, 5 d.f., $P = 0.03$) in the $G^2$ relative to the inpatient-only model when both 1-year inpatient and outpatient Charlson scores are included in the model as separate vectors (this corresponds to an increase in the $R^2$ analog from 0.038 to 0.046.). There is a further rise in $G^2$ of 9.4 points when the Charlson based on a 1-year lookback in the auxiliary claims is added (Fig. 1, $\Delta G^2 = 74.7 - 65.3 = 9.4$, 5 d.f., $P = 0.09$). A similar pattern of informativeness of different data sources as gauged by changes in the area under the ROC curve can also be noted in Table 2 and is shown visually in Fig. 2. The area under the ROC curve with demographics, primary diagnosis, and 1-year inpatient Charlson score was 0.702; with the addition of the 1-year outpatient and auxiliary claims, the area increased to 0.724 ($P = 0.02$).

As shown in Figs. 1 and 2 and in Table 2, a regular pattern was also found across the different tests for the informativeness of the second year of data for the different data sources. A second year of inpatient data was valuable. However, the use of the 366-to-730-day lookback within the alternative (ie, outpatient or auxiliary) data sources did not improve the performance of that Charlson score measure. More specifically, as shown in Fig. 1, there was a meaningful increase in the likelihood ratio $\chi^2$ statistic in the Cox regression models comparing a model with the 1-year inpatient

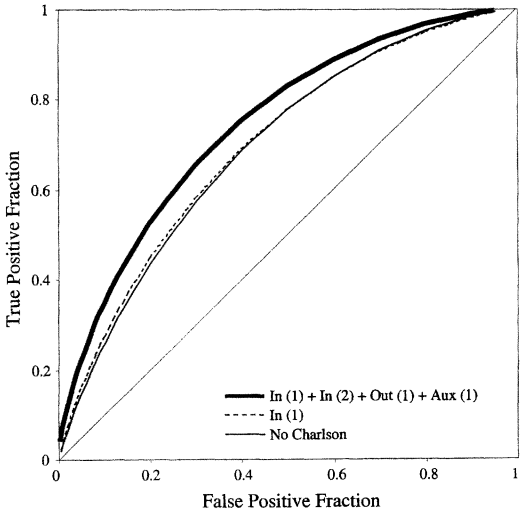


FIG. 2. ROC curves for predicting 365-day mortality from alternative data sources. Displayed are the ROC for three logistic regression models predicting death within 1-year of admission for the index hospitalization. All models control for patient demographics and primary diagnosis. In the model with multiple data sources for the Charlson score, each entered the regression as a separate vector. The area under the ROC curve without Charlson adjustment is 0.697, with 1-year inpatient claims based adjustment ("In [1]") is 0.702, and 2 distinct years of inpatient claims, 1 year of outpatient claims, and 1 year of auxiliary claims lead to an area under the ROC curve of 0.741.

lookback to a model with both the first and the second year of inpatient lookback (Fig. 1, $\Delta G^2$ = 63.4 - 53.1 = 10.3, 5 d.f., $P = 0.06$). However, there was no increase in $G^2$ for the addition of the second year of outpatient lookback to a model that already controlled for 1-year outpatient lookback, demographics, and primary diagnosis (Fig. 1, $\Delta G^2$ = 61.1 - 56.5 = 4.6, 5 d.f., $P = 0.47$). Similarly, the addition of the second year of auxiliary data did not significantly increase the $G^2$ of any model (eg, Fig. 1, $\Delta G^2$ = 90.1 - 83.8 = 6.3, 10 d.f., $P = 0.74$). This overall pattern was confirmed by inspecting the nonsignificant, individual coefficients for the additional data in the nested Cox regressions (data not shown) and by examining the changes in areas under the ROC curve presented in Table 2 ("Separate Vector" column). In summary, the addition of a second year of inpatient data to a model already containing 1 year of inpatient lookback produced a meaningful difference in the areas under the ROC curve (area under the ROC increased from 0.702 to 0.720; $P = 0.03$); however, the addition of a second year of outpatient data or of auxiliary data did not produce meaningful or statistically significant changes.

## Marginal Detection Efficacy of Data Sources

The marginal detection efficacy of each source of data for each of the 17 constituent diseases of the Charlson score is shown in Table 3. The table is read as follows: 27 patients were found to have myocardial infarction indicated as a diagnosis on an inpatient hospitalization claim for which the patient was discharged in the 365 days preceding the index hospitalization admission. An additional nine patients were found to have such a comorbidity when inspecting the outpatient claims for the same period. Twenty-nine additional patients were indicated to have such a comorbidity in the Auxiliary claims, which brought the total of patients with a Charlson Score contribution from myocardial infarction to 65. However, the relative proportion of new cases identified in each source varied across diagnoses; thus, the use of additional sources of data contributed to the Charlson score by detecting the constituent disease differentially.

## The Independent Value of Self-Reported Comorbidity

We also evaluated the contribution of self-reported comorbidity to the construction of Charlson comorbidity indices. When included as an undifferentiated data source in the single-vector Charlson Index, self-reported diagnoses had little or no value as compared with exclusive claims-based comorbidity detection. When included in the regressions as a separate vector of dummies, the addition of self-report data to a model containing the 1-year inpatient lookback increased $G^2$ by 13.4 points (Fig. 1, $\Delta G^2$ = 66.5 - 53.1 = 13.4, 5 d.f., $P = 0.02$; $R^2$ increased from 0.038 to 0.047). The addition of self-report data to a model containing separate vector Charlsons for 1-year inpatient, outpatient, and auxiliary claims, and a second-year inpatient lookback increased $G^2$ by 4.7 points (Fig. 1, $\Delta G^2$ = 88.5 - 83.8 = 4.7, 5 d.f., $P = 0.45$).

In the ROC analysis shown in Table 2, self-report data failed to significantly increase the area under the ROC curve versus regressions containing demographics, primary diagnosis, and either just 1-year inpatient-based Charlson or 1-year inpatient, 1-year outpatient, 1-year auxiliary, and a second-year of inpatient-based separate-vector Charlson scores. Conversely and more importantly, the claims-based Charlson scores did increase the area under the ROC curve relative to a model containing demographics, primary diagnosis, and a self-report-based Charlson score; the area increased to 0.743 from 0.704 ($P = 0.03$), with the addition of the Charlson based on a 1-year inpatient, 1-year outpatient, 1-year auxiliary, and a second-year of inpatient-based separate vector Charlson scores.

## Similarity of Pattern When Using Secondary Diagnoses From Index Hospitalization

When combined into a single Charlson score, the addition of secondary diagnoses from the index hospitalization itself significantly increases the predictive power of 1-year inpatient-only Charlson score (based on $G^2$ in Fig. 1 of 78.2 and 53.1, $R^2$ increases to 0.055 from 0.038). However, even better performance is achieved in a model that omits the retrospective inpatient data from the single Charlson score and uses only the secondary diagnoses (based on $G^2$ in Fig. 1 of 92.2 and 53.1, $R^2$ is 0.064 vs. 0.038).

TABLE 3. Marginal Detection of Constituent Comorbidities by Alternative Data Sources

| Disease | Weight | Year-1 Inpatient (detected patients) | Year-1 Outpatient But Not in Year-1 Inpatient (detected patients) | Year-1 Auxiliary But Not in Year-1 Inpatient or Outpatient (detected patients) |
|---|---|---|---|---|
| Myocardial infarction | 1 | 27 | 9 | 29 |
| Congestive heart failure | 1 | 71 | 65 | 142 |
| Peripheral vascular disease | 1 | 22 | 31 | 111 |
| Cerebrovascular disease | 1 | 35 | 49 | 121 |
| Dementia | 1 | 14 | 23 | 64 |
| Chronic pulmonary disease | 1 | 70 | 42 | 138 |
| Rheumatologic disease | 1 | 6 | 6 | 37 |
| Peptic ulcer disease | 1 | 8 | 11 | 28 |
| Liver disease (mild) | 1 | 0 | 3 | 3 |
| Diabetes (mild or moderate) | 1 | 34 | 65 | 150 |
| Diabetes (severe) | 2 | 20 | 10 | 60 |
| Hemiplagia or paraplegia | 2 | 12 | 12 | 13 |
| Renal disease | 2 | 7 | 15 | 14 |
| Malignancy | 2 | 14 | 72 | 92 |
| Liver disease (moderate or severe) | 3 | 0 | 0 | 2 |
| Metastatic solid tumor | 6 | 5 | 14 | 17 |
| AIDS | 6 | 0 | 0 | 1 |

Note: For each Charlson disease, the marginal number of patients who were found to have that disease by data source is shown; each column excludes any cases also noted to have that disease in a data source indicated in a column to its left. The original weights assigned by Charlson et al[1] and used in this study are provided for reference.

This pattern does not hold if the data are treated in the separate vector specification. In that case, a pattern similar to that observed in other separate-vector parameterizations occurs. The addition to the baseline model (controlling for demographics and the primary diagnosis of the index hospitalization) of a Charlson score based on secondary diagnoses from the index hospitalization raised likelihood ratio $\chi^2$ statistic by 49.7 points (Fig. 1, $\Delta G^2 = 92.2 - 42.5 = 49.7$, 5 d.f., $P < 0.001$). The further addition of the 1-year inpatient Charlson raised the $G^2$ an additional 4.3 points. (Fig. 1, $\Delta G^2 = 96.5 - 92.2 = 4.3$, 5 d.f., $P = 0.51$). The addition of other claims-based Charlson scores and/or the addition of self-report-based Charlson all also raised the $G^2$, although not by statistically significant amounts. Similarly, the addition of the three additional claims-based Charlson scores did increase the area under the ROC curve from 0.727 (for demographics plus primary diagnosis plus secondary-diagnosis-based Charlson plus 1-year inpatient Charlson) to 0.751, a statistically significant increase in predictive power ($P = 0.01$). However, there are interpretive difficulties in using any of these measures derived from the secondary diagnoses of the index hospitalization that are discussed later.

## Discussion

In a representative sample of Medicare beneficiaries, we examined the performance of Charlson scores based on alternative sources of data. Statistically and empirically significant improvements in the prediction of mortality can be obtained by incorporating alternative sources of data (particularly 2 years of inpatient lookback combined with 1 year of outpatient and auxiliary claims lookback) but only if indices derived from distinct sources of data are entered into the regression distinctly. Furthermore, we found that these improvements in explanatory power were largely true whether

one also controlled for Charlson scores based on self-reported health history and/or based on the secondary diagnoses from the claim for the index hospitalization.

Surprisingly, our results overall showed that the Charlson indices provided only modest improvements over simply controlling for the age, race, sex, and index hospitalization primary diagnosis of the patients. Among papers that report evaluations of comorbidity adjustment, similarly modest performance has been reported in some cases,[7,17,18] although not always.[25] In absolute magnitude, the areas under the ROC curve that we report are quite similar to those published previously using an inpatient-only Charlson score to predict in-hospital mortality among coronary artery bypass patients.[17] More generally, the Charlson score has typically been validated by demonstrating differences in utilization or mortality between score levels rather than by assessing its absolute increase in explanatory power.[1–4,14,19]

While we found that the explanatory power of the Charlson score could be augmented by the use of survey-derived self-report of health history, as has been suggested by previous work using the SF-36 and other comorbidity adjustment schemes,[25,36–38] we also found that an index based on inpatient claims data alone had approximately the same explanatory power as an index based on survey-derived data alone. Although survey-derived data are often not available, when they are, they seem to tap somewhat distinct "health" information as compared with the inpatient claims; self-report of health history and outpatient/auxiliary claims may be substitutes to each other. These conclusions, however, are particularly dependent on the self-reported health history instrument available. The instrument in the MCBS was not optimized for the development of Charlson scores and superior performance would probably be obtained with a more focused survey.[36]

Finally, our results demonstrated that the use of earlier claims can significantly augment risk adjustment using the secondary diagnoses of the index hospitalization. It is well known that there are important conceptual difficulties with the use of secondary diagnoses from the index hospitalization to adjust for the prehospitalization level of comorbidity in a patient population; in particular, it is impossible to assess whether the secondary diagnoses from the index hospitalization represent true pre-existing comorbidities that complicated the patient's care (and, hence, are appropriate for risk adjustment), or rather if they represent the result of complications and suboptimal treatment of a patient (and, hence, should be considered outcomes of care, not comorbidities).[5] Moreover, the fraction of time that any individual diagnosis is a complication rather than a comorbidity may vary as a function of both the institutions and procedures under study.[46] For the purposes of this article, we do not need to take a position on this methodological debate but merely note that if one chooses to proceed with risk adjustment using secondary diagnoses, one can still improve the accuracy of the model by using diverse prior claims data sources, as well (if the information from different data sources is incorporated distinctly).

This work is not without its limitations. First and foremost, we have looked only at ways in which the conventional ICD-9-CM-based versions of the Charlson score may be operationalized in the claims. Our results are subject to all the significant and well-known limitations of the Charlson score implemented in administrative records.[3,26,28,46,47] Second, we used a representative sample of the elderly and looked at their hospitalizations and consequent mortality; thus, our population is relatively more healthy than a representative sample of hospitalizations. Different performance characteristics might be found in different subpopulations. Our use of the MCBS allowed us to look at many different sources of data, including self-reported health history; however, the small size of the data set limited our ability to perform detailed analyses on restricted subpopulations. Third, our ability to generate a Charlson score from the self-reported health history is obviously dependent on the particular questionnaire that was used. Fourth, we have not performed an exhaustive search of alternative specifications (for example, the use of a quadratic continuous Charlson score instead of our multiple indicator variable approach) nor for alternative outcomes (such as inpatient mortality, total resource use, or length of stay); naturally, alternative data sources might perform differently when validated against different outcomes.

Our data do confirm the following: (1) that the Charlson comorbidity index, in conjunction with basic demographics, does have explanatory power to predict mortality following hospitalization, and (2) that the simple use of additional, readily available claims data sources can significantly enhance that explanatory power.

## Acknowledgment

The authors thank Marshall Chin for his helpful comments and Kim Thomas for administrative support.

## References

1. **Charlson ME, Pompei P, Ales KL, MacKenzie CR.** A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation. J Chronic Dis 1987;40:373–383.

2. **Deyo RA, Cherkin DC, Ciol MA.** Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. J Clin Epidemiol 1992;45:613–619.

3. **Romano PS, Roos LL, Jollis JG.** Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: Differing perspectives. J Clin Epidemiol 1993;46:1075–1079.

4. **D'Hoore W, Sicotte C, Tilquin C.** Risk-adjustment in outcome assessment: The Charlson comorbidity index. Methods Inf Med 1993;32:382–387.

5. **Iezzoni LI.** Risk adjustment for measuring health outcomes. Chicago: Health Administration Press, 1997:73–78.

6. **Elixhauser A, Steiner C, Harris DR, Coffey RM.** Comorbidity measures for use with administrative data. Med Care 1998;36:8–27.

7. **Schwartz M. Iezzoni LI, Moskowitz MA, Ash AS, Sawitz E.** The importance of comorbidities in explaining differences in patient costs. Med Care 1996;34:767–782.

8. **Iezzoni LI, Daley J, Heeren T, et al.** Using administrative data to screen hospitals for high complication rates. Inquiry 1994;31:40–55.

9. **Kuykendall DH, Ashton CM, Johnson ML, Geraci JM.** Identifying complications and low provider adherence to normative practice using administrative data. Health Serv Res 1995;30:531–554.

10. **D'Hoore W, Bouckaert A, Tilquin C.** Practical considerations on the use of the Charlson comorbidity index with administrative data bases. J Clin Epidemiol 1996;49:1429–1433.

11. **Christakis NA, Escarce JJ.** Survival of Medicare patients following enrollment in hospice programs. N Engl J Med 1996;335:172–178.

12. **Iwashyna TJ, Zhang JX, Lauderdale DS, Christakis NA.** A method for identifying married couples in the Medicare claims data: Mortality, morbidity and health care utilization among the elderly. Demography 1998;35:413–419.

13. **Roos NP, Wennberg JE, Malenka DJ, et al.** Mortality and reoperation after open and transurethral resection of the prostate for benign prostatic hyperplasia. N Engl J Med 1989:320:1120–1124.

14. **Romano PS, Roos LL, Jollis JG.** Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. J Clin Epidemiol 1993;46:1085–1090.

15. **Roos LL, Sharp SM, Cohen MM.** Comparing clinical information with claims data: Some similarities and differences. J Clin Epidemiol 1991;44:881–888.

16. **Hughes JS, Iezzoni LI, Daley J, Greenberg L.** How severity measures rate hospitalized patients. J Gen Intern Med 1996;11:303–311.

17. **Ghali WA, Hall RE, Rosen AK, Ash AS, Moskowitz MA.** Searching for an improved clinical comorbidity index for use with ICD-9-CM administrative data. J Clin Epidemiol 1996;49:273–278.

18. **Kieszak SM, Flanders WD, Kosinski AS, Shipp CC, Karp H.** A comparison of the Charlson comorbidity index derived from the medical record data and administrative data. J Clin Epidemiol 1999;52:137–142.

19. **Librero J, Peiro S, Ordinana R.** Chronic comorbidity and outcomes of hospital care: Length of stay, mortality, and readmission at 30 and 365 days. J Clin Epidemiol 1999;52:171–179.

20. **DesHarnais SI, McMahon Jr. LF, Wroblewski RT, Hogan AJ.** Measuring hospital performance: The development and validation of risk-adjusted indexes of mortality, readmissions, and complications. Med Care 1990;28:1127–1141.

21. **Brailer DJ, Kroch E, Pauly MV, Hunag J.** Comorbidity-adjusted complication risk. Med Care 1996;34:490–505.

22. **Weiner JP, Starfield BH, Steinwachs DM, Mumford LM.** Development and application of a population-oriented measure of ambulatory care case-mix. Med Care 1991;29:452–472.

23. **Starfield B, Weiner J, Mumford L, Steinwachs D.** Ambulatory Care Groups: A categorization of diagnoses for research and management. Health Serv Res 1991;26:53–74.

24. **Fedson DS, Wajda A, Nicol JP, Hammond GW, Kaiser DL, Roos LL.** Clinical effectiveness of influenze vaccination in Manitoba. JAMA 1993;270:1956–1961.

25. **Fowles JB, Weiner JP, Knutson D, Folwer E, Tucker AM, Ireland M.** Taking health status into account when setting capitation rates: A comparison of risk-adjustment methods. JAMA 1996;276:1316–1321.

26. **Iezzoni LI, Foley SM, Daley J, Hughes J, Fisher ES, Heeren T.** Comorbidities, complications, and coding bias: Does the number of diagnosis codes matter in predicting in-hospital mortality? JAMA 1992;267:2197–2203.

27. **Green J, Wintfeld N.** How accurate are hospital discharge data for evaluating effectiveness of care? Med Care 1993;31:719–731.

28. **Romano PS, Mark DH.** Bias in the coding of hospital discharge data and its implications for quality assessment. Med Care 1994;32:81–90.

29. **Fisher ES, Whaley FS, Krushat M, et al.** The accuracy of Medicare's hospital claims data: Progress has been made, but problems remain. Am J Pub Health 1992;82:243–248.

30. **Fowles JB, Lawthers AG, Weiner JP, Gernick DW, Petrie DS, Palmer RH.** Agreement between physicians' office records and Medicare Part B claims data. Health Care Financ Rev 1995;16:189–199.

31. **Lauderdale DS, Goldberg J.** The expanded racial and ethnic codes in the Medicare data files: Their completeness of coverage and accuracy. Am J Pub Health 1996;86:712–716.

32. **Deyo RA.** Adapting a clinical comorbidity index for use with ICD-9-CM administrative data: A response. J Clin Epidemiol 1993;46:1081–1082.

33. **McBean AM, Warren JL, Babish JD.** Measuring the incidence of cancer in elderly Americans using Medicare claims data. Cancer 1994;73:2417–2425.

34. **McClellan M, Roghmann H, Schilling P.** The validity of Medicare claims for studies of cancer incidence: Breast, prostate, and lung cancer in the elderly. Bethesda, MD: SEER-Medicare Data Users Workshop, June 24, 1998.

35. **Klabunde C.** Assessment of comorbidity using claims data. Bethesda, MD: SEER-Medicare Data Users Workshop, June 24, 1998.

36. **Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW.** Can comorbidity be measured by questionnaire rather than medical record review? Med Care 1996;34:73–84.

37. **Muhajarine N, Mustard C, Roos LL, Young TK, Gelskey DE.** Comparison of survey and physician claims data for detecting hypertension. J Clin Epidemiol 1997;50:711–718.

38. **Robinson JR, Young TK, Roos LL, Gelskey De.** Estimating the burden of disease: Comparing administrative sources of data and self-report. Med Care 1997;35:932–947.

39. **Adler GS.** A profile of the Medicare current beneficiary survey. Health Care Financ Rev 1994;15:153–163.

40. **Collins JG.** Prevalence of selected chronic conditions: United States, 1990–1992. National Center for Health Statistics. Vital Health Stat 1997;10:(194).

41. **Allison PD.** Survival analysis using the SAS System: A practical guide. Cary, NC: SAS Institute, 1995.

42. **Magee L.** R-Squared measures based on Wald and likelihood ratio joint significance tests. Am Statistician 1990;44:250–253.

43. **Metz CE.** Basic principles of ROC Analysis. Sem Nuclear Med 1978;8:283–298.

44. **Metz CE, Wang P-L, Kronman HB.** A new approach for testing the significance of differences between ROC curves measured from correlated data. In: Deconinck F, ed. Information Processing in Medical Imaging. The Hague: Martinus Nijhoff, 1984:432–445.

45. **Metz CE, Herman BA, Roe CA.** Statistical comparisons of two ROC curve estimates obtained from partially-paired data sets. Med Decis Making 1998;18:110.

46. **Roos LL, Stranc L, James RC, Li J.** Complications, comorbidities, and mortality: Improving classification and prediction. Health Serv Res 1997;32:229–238.

47. **Iezzoni LI.** Assessing quality using administrative data. Ann Intern Med 1997;127:666–674.